

Klieme, Eckhard; Leutner, Detlev

Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG

Zeitschrift für Pädagogik 52 (2006) 6, S. 876-903



Quellenangabe/ Reference:

Klieme, Eckhard; Leutner, Detlev: Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG - In: Zeitschrift für Pädagogik 52 (2006) 6, S. 876-903 - URN: urn:nbn:de:0111-opus-44936 - DOI: 10.25656/01:4493

<https://nbn-resolving.org/urn:nbn:de:0111-opus-44936>

<https://doi.org/10.25656/01:4493>

in Kooperation mit / in cooperation with:

BELTZ

<http://www.beltz.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Inhaltsverzeichnis

Thementeil: Videogestützte Unterrichtsforschung

Eckhard Klieme

Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einleitung in den Thementeil 765

Christine Pauli/Kurt Reusser

Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung 774

Tina Seidel/Manfred Prenzel/Rolf Rimmel/Inger Marie Dalehefte/

Constanze Herweg/Mareike Kobarg/Katharina Schwindt
Blicke auf den Physikunterricht. Ergebnisse der IPN Videostudie 798

Katrin Rakoczy

Motivationsunterstützung im Mathematikunterricht: Zur Bedeutung von Unterrichtsmerkmalen für die Wahrnehmung der Schülerinnen und Schüler 822

Deutscher Bildungsserver

Linktipps zum Thema Unterrichtsqualität 844

Allgemeiner Teil

Katharina Maag Merki

Risikosubstanzenkonsum und somatische Beschwerden. Ergebnisse einer Längsschnittstudie bei Schülerinnen und Schülern der gymnasialen Oberstufe ... 855

Dokumentation 1

Eckhard Klieme/Detlev Leutner

Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen 876

Besprechungen

Norbert Ricken

Klaus Prange: Die Zeigestruktur der Erziehung. Grundriss der Operativen Pädagogik	904
---	-----

Monika A. Vernooij

Ada Sasse: Sonderschüler und Sonderschulen im ländlichen Bereich	908
--	-----

Günther Deegener

Barbara Kavemann/Ulrike Kreyssig (Hrsg.): Handbuch Kinder und häusliche Gewalt	912
--	-----

Dokumentation 2

Pädagogische Neuerscheinungen	916
-------------------------------------	-----

Beilagenhinweis:

Dieser Ausgabe der ZfPäd liegt das Jahresarhaltsverzeichnis 2006 bei.

Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen

Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG¹

Anlass und Ziel des Schwerpunktprogramms

Gesellschaftlicher Wohlstand, soziale Kohäsion und Entwicklungschancen einer Gesellschaft hängen in entscheidendem Maße vom Bildungsstand ihrer Mitglieder ab. Die Fachdiskussion in der Bildungsforschung, aber z.B. auch in der Wirtschaftswissenschaft (Heckman, 2000) macht deutlich, dass man den Stand der Humanressourcen nicht allein nach den Zertifikaten bewerten kann, die in Institutionen der formalen Bildung erworben werden. Die „Produkte“ der Bildungs- und Ausbildungsprozesse selbst, d.h. das erworbene Wissen, die ausgebildeten Fähigkeiten, Fertigkeiten, Einstellungen und Bereitschaften und die verfügbaren Qualifikationen, sind daher unmittelbar in den Blick geraten.

Zunächst in der Weiterbildung und der beruflichen Bildung, sodann auch in der allgemeinen Schulbildung und der Hochschulbildung wurde der Begriff der *Kompetenzen* zum Ankerpunkt der Diskussion über den „Output“ von Bildung und Ausbildung (zur Definition vgl. den folgenden Abschnitt). Beispielsweise erfasste der PISA-Test zur Lesekompetenz (Artelt, Stanat, Schneider & Schiefele, 2001; Schaffner u.a., 2004), wie gut Jugendliche in alltagsnahen Kontexten mit narrativen Texten, aber auch mit so genannten diskontinuierlichen Texten (Graphiken, Tabellen) umgehen können, und in beiden Teilbereichen (Dimensionen) wurden Kompetenzniveaus unterschieden, die von der Entnahme einzelner Informationselemente bis zur kritischen Reflexion von Textaussagen reichen. Mit dem Kompetenzbegriff kamen auch komplexere Anforderungen in den Blick wie z.B. die mathematische Modellierung von Alltagssituationen, die Planung und Auswertung von naturwissenschaftlichen Experimenten, Selbstregulation und Metakognition beim Lösen fachspezifischer Probleme.

Internationale Vergleichsuntersuchungen am Ende der Pflichtschulzeit (z.B. TIMSS, PISA) und bei Erwachsenen (z.B. IALS, ALL) haben große Aufmerksamkeit gefunden, weil sie die Diskrepanz zwischen den Zielen und Ansprüchen des Bildungssystems einerseits und den tatsächlich erreichten Kompetenzen andererseits sichtbar machen

1 Leicht überarbeitete Fassung des Antrags bei der Deutschen Forschungsgemeinschaft. Die Autoren danken Johannes Hartig (DIPF) und den Teilnehmern eines DFG-Rundgesprächs für wertvolle Anregungen.

konnten. Als Ergebnis wurden – jedenfalls im Schulwesen – weit reichende Maßnahmen in Angriff genommen wie z.B. gezielte Unterrichtsreformen für Mathematik und Naturwissenschaften (die bei der jüngsten PISA-Untersuchung bereits Wirkung zu zeigen schienen; vgl. Prenzel et al., 2004, 2005) oder Initiativen zur individuellen Förderung der Lesekompetenz. Ein Kernelement der neuen Reformagenda – das sie von früheren Bildungsreformen unterscheidet – ist der Aufbau von Evaluationsagenturen, die Lernergebnisse sowohl auf der Ebene einzelner Klassen und Schulen als auch auf der Ebene der Bildungssysteme ermitteln und an die Akteure zurückmelden. Schulpolitik und Schulpraxis verändern sich dadurch nachhaltig in Richtung auf *evidence-based policy and practice* (Slavin, 2002; vgl. auch Ditton, 2000). In anderen Bildungsbereichen (vorschulische Bildung und Betreuung, berufliche Aus- und Weiterbildung, Hochschulen) bahnen sich analoge Entwicklungen an. Sowohl zur Evaluation der Bildungseinrichtungen als auch zur Begründung pädagogischer Maßnahmen im Einzelfall müssen Lernvoraussetzungen und Lernergebnissen differenziert erfasst werden. Beispielsweise zielen neue Bachelor/Master-Studiengänge auf die Vermittlung spezifischer Kompetenzen, die prinzipiell überprüfbar sein sollen, und auch bei der Zulassung zum Studium werden in wachsender Zahl eignungsdiagnostische Verfahren verwendet (vgl. die diesbezüglichen Empfehlungen der Dt. Gesellschaft für Psychologie; Gold & Souvignier, im Druck; Köller, 2004).

Der Messung von Kompetenzen kommt demnach eine Schlüsselfunktion für die Optimierung von Bildungsprozessen und für die Weiterentwicklung des Bildungswesens zu. Von der pädagogisch-psychologischen Forschung wird erwartet, dass sie entsprechende Messverfahren bereitstellt als Basis für Förder-, Platzierungs- und Auswahlentscheidungen, für Benotung und Zertifizierung von Lernenden, aber auch für die Evaluation von pädagogischen Maßnahmen und Institutionen sowie die laufende Beobachtung der Qualität von Bildungssystemen und ihrer gesellschaftlichen Wirkungen. Mit dem Schwerpunktprogramm (SPP) sollen die wissenschaftlichen Voraussetzungen zur Erfüllung dieser Aufgaben maßgeblich verbessert werden. Wie anspruchsvoll die Erfassung von Lernvoraussetzungen und -ergebnissen theoretisch und methodisch ist, wird in der Bildungspraxis und -politik häufig nicht wahrgenommen. Kompetenzmessung muss die Binnenstruktur der Kompetenzen, d.h. die Teilfähigkeiten (z.B. Umgang mit narrativen Texten einerseits, mit diskontinuierlichen Texten andererseits) ebenso wie die Niveaustufungen differenziert darstellen; sie soll zudem Veränderungen im Lern- und Entwicklungsprozess abbilden können. Die *Messverfahren* bedürfen dazu der Verankerung in empirisch geprüften kognitionspsychologischen bzw. fachdidaktischen *Kompetenzmodellen*, die Struktur, Graduierung und Entwicklungsverläufe der Kompetenzen abbilden. Die Kompetenzmessung muss sich darüber hinaus auf *psychometrische Modelle* stützen, die gewissermaßen zwischen Messoperation und Kompetenzmodell vermitteln, d.h. begründen, wie ein Messergebnis (z.B. Anzahl und Art der gelösten Aufgaben) im Sinne des Kompetenzmodells zu interpretieren ist.

Modelle und Verfahren sind im SPP ausnahmslos auf konkrete Lern- und Handlungsbereiche bezogen, die für Bildung und Ausbildung relevant sind. Es geht also z.B. nicht um schlussfolgerndes Denken als abstraktes psychologisches Konstrukt, sondern –

im Beispiel bleibend – um kaufmännische Kompetenz von Auszubildenden im Sinne der Fähigkeit, Abläufe in einem simulierten Betrieb zu verstehen. Die Ausgestaltung der Modelle und Messverfahren hängt in solchen Kontexten davon ab, welche Fragen sie beantworten bzw. welche Entscheidungen sie begründen sollen. Ein Verfahren, das für gewisse Zwecke angemessen ist (z.B. Lernenden eine unmittelbare Rückmeldung zu geben), kann für andere Zwecke (z.B. die vergleichende Evaluation von Bildungseinrichtungen) völlig ungeeignet sein – etwa weil im ersten Fall ein detaillierteres Kompetenzmodell erforderlich ist als im zweiten und weil die Präzision der Messung einmal auf der Individualebene, ein andermal auf der Aggregatebene erforderlich ist. Die Vermischung beider Zwecke ist mit vielfältigen Problemen behaftet, wie Erfahrungen vor allem aus den USA zeigen (Cheng, Watanabe & Curtis, 2004; Fuhrman & Elmore, 2004). Zum SPP sollen daher auch Forschungsvorhaben gehören, sie sich mit der *Nutzung* von Kompetenzmodellen in unterschiedlichen pädagogischen Entscheidungssituationen befassen.

Das SPP greift eine aktuelle Thematik mit neuer Qualität auf. Gute Voraussetzungen wurden in den letzten Jahren durch Entwicklungen im Bereich psychometrischer Messmodelle sowie in den für empirische Messverfahren zur Verfügung stehenden Technologien geschaffen. Die systematische Verbindung zwischen kognitiven Kompetenzmodellen, psychometrischen Modellen und Messverfahren in unterschiedlichen pädagogischen Kontexten aber ist auch international als wissenschaftlich noch nicht hinreichend gelöstes, wenngleich praktisch sehr drängendes Problem erkannt worden. So hat das vom US-amerikanischen *National Research Council* eingesetzte „Committee on the Foundations of Assessment“ (Pellegrino, Chudowsky & Glaser, 2001) multidisziplinäre Forschungsaktivitäten mit drei Facetten gefordert: „(1) development of cognitive models of learning that can serve as the basis for assessment design, (2) research on new statistical measurement models and their applicability, (3) research on assessment design“ (S. 284).

Die Bildungsforschung in Deutschland ist aufgrund ihrer inzwischen starken Beteiligung an internationalen Assessment-Projekten (s.u.) und aufgrund hervorragender Präsenz in relevanten psychologischen Forschungsfeldern gut vorbereitet, diese Fragen auf international konkurrenzfähigem Niveau zu bearbeiten. Als ein auf Grundlagenforschung fokussierendes koordiniertes Forschungsprogramm wäre das SPP nach gegenwärtigem Stand der Erkenntnis sogar international einzigartig. Es würde Deutschland nicht nur international anschlussfähig machen, sondern hätte auch gute Chancen, Deutschland international zu einer führenden Rolle zu verhelfen. Auf nationaler Ebene kann das SPP darüber hinaus eine neue Qualität von *Kooperation* beanspruchen, indem forschungsproduktive Wissenschaftler, die z.T. auch in angewandten Diagnostik- und Assessment Projekten arbeiten, im Hinblick auf Grundlagenforschung in diesem Bereich zusammengeführt und gestärkt werden.

Das angestrebte Schwerpunktprogramm soll dementsprechend als integratives Forschungsprogramm kognitiv orientierte Experten auf den Gebieten einzelner bereichsspezifischer Kompetenzen mit Experten auf dem Gebiet des Messens und Testens zusammenbringen. Ziel ist es, Kompetenzstruktur- und Kompetenzentwicklungsmodelle zu erarbeiten und

empirisch zu prüfen, anhand derer sich valide und faire Messinstrumente auf zwei Ebenen konstruieren lassen: zum einen auf der Ebene der Förderung individueller Lernprozesse (im Sinne einer individuellen „Kompetenzdiagnostik“), zum anderen auf der Ebene von Untersuchungen zum Monitoring von Bildungsinstitutionen und Bildungssystemen („Assessment“). Abgerundet wird das Programm durch Forschung zur Nutzung von kompetenzbezogenen Messinstrumenten in unterschiedlichen pädagogischen Entscheidungskontexten.

Wissenschaftliches Programm

Kompetenzbegriff, Domänen und Kompetenzmodelle

Der Kompetenzbegriff ist zentral für empirische Studien, die sich mit der Entwicklung von Humanressourcen und der Produktivität des Bildungswesens befassen. Obwohl schon vor Jahrzehnten eingeführt, ist er erst in den letzten Jahren zum Gegenstand intensiver Diskussion in der Psychologie und ihren Nachbardisziplinen geworden (z.B. Csapó 2004; Klieme, Funke, Leutner, Reimann & Wirth, 2001; Sternberg & Grigorenko, 2003; Rychen & Salganik, 2001, 2003; Weinert, 2001b). Nicht zuletzt greift die Forschung diesen Begriff auf, um veränderte Anforderungen der Lebens- und Arbeitswelt sowie damit zusammenhängende Bildungsziele zu charakterisieren. Für das SPP definieren wir Kompetenzen als *kontextspezifische kognitive Leistungsdispositionen*, die sich funktional auf Situationen und Anforderungen in bestimmten *Domänen* beziehen.

Das wesentliche Charakteristikum des hier verwendeten Kompetenzbegriffs ist die *Kontextabhängigkeit*. So wurde der Kompetenzbegriff in der Psychologie als Gegenbegriff zur klassischen Intelligenzforschung eingeführt, die generalisierte, kontextunabhängige, nur begrenzt erlernbare kognitive Dispositionen untersucht (z.B. McClelland, 1973; „Testing for competence rather than for ‚intelligence“). Kompetenz bezieht sich immer darauf, Anforderungen in spezifischen Situationen bewältigen zu können. So definiert schon White (1959, S. 317) Kompetenz als „effective interaction (of the individual) with the environment“; besonders prägnant bezeichnen Connell et al (2003, S. 142) Kompetenzen als „realized abilities“.

Fragestellungen im SPP sollen sich auf Kompetenzen in unterschiedlichen Lernfeldern (z.B. Lesen) oder fachbezogenen Leistungsbereichen (z.B. mathematisches Modellieren) beziehen; in der Tradition der psychologischen Expertiseforschung sprechen wir von *Domänen*. Eine besondere Stärke der Bildungsforschung in Deutschland sind Arbeiten zur Lesekompetenz (Artelt et al., 2001; Artelt, Schiefele & Schneider, 2001; Dutke, 2000; Grube & Hasselhorn, im Druck; Herzmann, Möller & Schiefele, 2004; Richter & Christmann, 2002; Schnotz & Dutke, 2004; Schwippert, Bos & Lankes, 2004; Souvignier, Küppers & Gold, 2003; Sparka & Gräsel, im Druck), aber auch zu Kompetenzen in mathematischen (Blum et al., 2004; Bruder, Perels, Gürtler & Schmitz, 2002; Kaiser, Hino & Knipping, 2005; Klieme & Baumert, 2001) und naturwissenschaftlichen (Fischer et al., im Druck; Klee, Sandmann & Vogt, 2005; Labudde et al., 2000; Möller, 2001; Prenzel et al., 2001; Rost et al. 2003; Schecker, 1999; Sumfleth, 2002) Domänen und zur Prob-

lemlösekompetenz (Baumert, Evans & Geiser, 1998; Frensch & Funke, 1995; Funke, 2003; Klieme, 2004; Klieme, Leutner & Wirth, 2005; Leutner, Klieme, Meyer & Wirth, 2004, 2005; Perels, Gurtler & Schmitz, 2005; Renkl & Atkinson, 2003; Wirth & Klieme, 2003). Das Programm soll an diese Stärken anknüpfen und neue Ansätze zur Messung von Kompetenzen entwickeln, die auf kognitiven Prozessmodellen aufbauen und beispielsweise die bei PISA derzeit genutzten Stufenbeschreibungen (z.B. der Lesekompetenz) ergänzen. Das Programm soll ferner Komponenten beruflicher und professioneller Expertise ansprechen wie z.B. kaufmännisches Entscheidungshandeln oder das Organisieren von Informationen in technischen Steuerungsprozessen (Breuer & Satish, 2004). Auch in diesen Bereichen wurden in Deutschland bereits qualitativ hochwertige Vorarbeiten geleistet (Boshuizen, Bromme & Gruber, 2004; Gruber, Harteis, Heid & Meier, 2004). Berufliche Kompetenzen schließen – wiederum im Sinne der Expertiseforschung (Gruber, 2004) – auch wohl definierte Aspekte der professionellen Kompetenz von Lehrpersonen (z.B. Hartinger & Fölling-Albers, 2004; Blömeke, 2003) wie z.B. diagnostische Kompetenz und Teilbereiche des *pedagogical content knowledge* (Weinert, Helmke & Schrader, 1992) mit ein.

Im Sinne einer inhaltlichen Fokussierung des SPP beschränkt sich der hier verwendete Kompetenzbegriff auf *kognitive* Dispositionen. Weinert (2001a) geht in seiner einflussreichen Expertise zum Kompetenzbegriff hingegen über den kognitiven Bereich hinaus und diskutiert auch so genannte Handlungskompetenzen, die motivationale Orientierungen, Einstellungen, Tendenzen und Erwartungen einschließen. Allerdings schlägt Weinert an derselben Stelle vor, in empirischen Untersuchungsdesigns kognitive und motivationale Tendenzen getrennt zu erfassen, gerade weil nur so ihre Wechselwirkung analytisch dargestellt werden kann. Die Trennung kognitiver und motivationaler Faktoren soll für das geplante Schwerpunktprogramm übernommen werden.

Ein für die Bildungsforschung wichtiges Merkmal des Kompetenzkonzepts ist schließlich, dass Kompetenzen durch Erfahrung und Lernen in relevanten Anforderungssituationen erworben sowie durch äußere Interventionen beeinflusst werden können (z.B. Baumert, Stanat & Demmrich, 2001; Hartig & Klieme, im Druck; Simonton, 2003). Im Unterschied hierzu sind kognitive Grundfunktionen in wesentlich geringerem Maße erlernbar und trainierbar (Weinert, 2001a). Bei der Formulierung von Kompetenzmodellen kann es eine wichtige Frage sein, die Beziehungen und Grenzen von spezifischen Kompetenzen und kognitiven Grundfunktionen einzubeziehen und auch empirisch zu untersuchen.

Die zentrale Aufgabe der Forschung besteht in der Entwicklung von Modellen der Struktur, Stufung und Entwicklung von Kompetenzen, die kognitionspsychologisch fundiert sind und mit avancierten psychometrischen Techniken verbunden werden können. Der Anspruch, kontextspezifische, erlernbare und vermittelbare Leistungsdispositionen zu erfassen, die für die Bewältigung von Anforderungen in konkreten Domänen erforderlich sind, stellt eine besondere Herausforderung dar. Die psychologische Grundlagenforschung und die Psychometrie sind diesem Anspruch bislang noch keineswegs gerecht geworden. Sie müssen ihre Modelle domänenspezifisch kontextualisieren und sind dabei auf eine Zusammenarbeit sowohl mit Erziehungswissenschaftlern

als auch mit Domänenexperten, etwa aus den Fachdidaktiken, angewiesen. Diese interdisziplinäre Kooperation soll das Schwerpunktprogramm leisten.

Erfassung von Kompetenzen im Bildungswesen: Diagnostik und Assessment

Thema des geplanten SPP ist die Modellierung und Erfassung von Kompetenzen im Bildungsbereich. Aus einer Nutzenperspektive sollen diese Kompetenzmodelle und Messinstrumente es zukünftig erlauben, bildungsbezogene Entscheidungen über Einzelpersonen besser als bisher vorzubereiten (z.B. hinsichtlich der Zuordnung zu einem bestimmten Bildungsgang oder der Vergabe eines bestimmten Bildungsabschlusses, aber auch im Hinblick auf Förderangebote; vgl. Holling, Vock & Wittmann, 2001). Solche Erkenntnisbemühungen im Hinblick auf individuelle Kompetenzen gehören zum Arbeitsfeld der pädagogischen bzw. pädagogisch-psychologischen Diagnostik (Klauer 1978; vgl. Leutner, 1992, 2001b; Wild & Krapp, 2001). Kompetenzdiagnostik im Bildungsbereich fokussiert dabei nicht auf Erziehungsprozesse allgemein, sondern auf Bildungsprozesse im Sinne formaler, institutionalisierter Lernprozesse, und sie bezieht sich immer auf Lern- und Handlungsbereiche (Domänen), die durch Bildungsziele vorstrukturiert sind.

Im Englischen spricht man etwas allgemeiner von *Educational Measurement* oder *Testing* und meint damit „a process by which educators use students' responses to specially created or naturally occurring stimuli to draw inferences about the students' knowledge and skills" (Pellegrino et al., 2001, S. 20). Der Begriff *Educational Assessment*, der mitunter synonym mit *Educational Measurement* und *Testing* verwendet wird, soll im Folgenden für Erhebungen reserviert werden, bei denen es nicht um individuelle Entscheidungen geht, sondern um eine Bilanzierung von Lernergebnissen auf Klassen-, Schul- und Systemebene. Hierzu gehören sowohl *classroom-based assessments* als auch *large-scale assessments* wie etwa das US-amerikanische *National Assessment of Educational Progress* (NAEP) oder PISA.

Diagnostik und *Assessment* sind schließlich begrifflich abzugrenzen von *Evaluation*. Bei *Evaluation* (vgl. Wottawa, 2001) geht es um eine Beurteilung von *Maßnahmen* hinsichtlich definierter *Ziele*. Im Bildungsbereich bedient sich *Evaluation* häufig derselben Konzepte und Verfahren, wie sie für individualdiagnostische Zwecke entwickelt wurden. Die Vermischung unterschiedlicher Erhebungszwecke innerhalb einer Erhebung ist jedoch problematisch, weil prinzipiell unterschiedliche Ansprüche an die Breite oder die Genauigkeit der Messung gestellt werden müssen.

Werden Differenzierungen nach Domänen und Erhebungszwecken nicht berücksichtigt, kann es zu oberflächlichen oder atomistischen Urteilen und zur Fehlsteuerung von Bildungsprozessen kommen. Die lern- bzw. lehrzielbezogene Leistungsmessung der 1970er-Jahre ist nicht zuletzt auch daran gescheitert, dass sie den Anspruch einer engen Verbindung von kognitiver Modellierung, Psychometrie und diagnostischen bzw. Assessment-Konzepten nicht einlösen konnte. Entwicklungen in den betroffenen Disziplinen ermöglichen jetzt eine neue Form von Diagnostik und Assessment, die auf intensi-

ver Grundlagenforschung beruht. Pellegrino et al. (2001) bringen die Herausforderung auf den Punkt: „Much hard work remains to focus psychometric model building on the critical features of models of cognition and learning and on observations that reveal meaningful cognitive processes in a particular domain. (...) Therefore, having a broad array of models available does not mean that the measurement model problem has been solved. The long-standing tradition of leaving scientists, educators, task designers, and psychometricians each to their own realms represents perhaps the most serious barrier to progress” (S. 6).

Zentrale Fragestellungen des Forschungsprogramms und Stand der Forschung

Die zentralen Fragestellungen des angestrebten SPP gliedern sich in vier aufeinander aufbauende Bereiche (Abbildung 1). Im Kern steht die Entwicklung theoretischer Kompetenzmodelle (Bereich 1), die durch psychometrische Messmodelle ergänzt werden (Bereich 2), die wiederum in die Konstruktion von Verfahren zur empirischen Erfassung von Kompetenzen münden (Bereich 3). Fragen zur Nutzung von diagnostischen Informationen (Bereich 4) runden das wissenschaftliche Programm ab. Der Bereich 3 betrifft keineswegs triviale Aufgaben, weil die domänenspezifische Kontextualisierung von Messverfahren in vielen Fällen neue Konzepte für Aufgaben und Anforderungen erfordert, z.B. den Einsatz neuer Technologien für computer- oder simulationsbasiertes Testen. Fragen zum Bereich 4 schließlich stehen zwar nicht im Zentrum des Programms, sind aber dennoch zu stellen und zu beantworten, um z.B. die Validität von

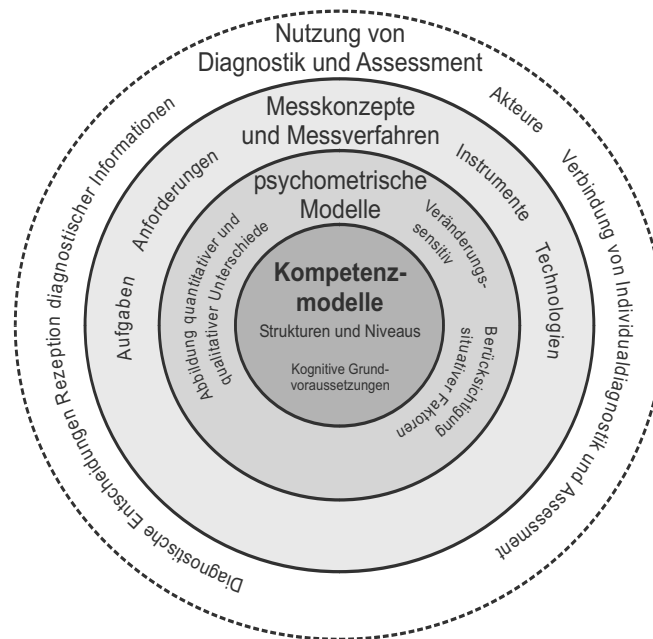


Abb. 1: **Arbeitsbereiche des Schwerpunktprogramms**

kompetenzbezogenen Urteilen beurteilen zu können. Im Bereich 4 sind insbesondere primär erziehungswissenschaftliche Studien zur diagnostischen Kompetenz von Lehrpersonen erforderlich.

In den folgenden Abschnitten werden die mit den Arbeitsbereichen des Programms verbundenen Fragestellungen ausgeführt, und der Stand der Forschung wird skizziert.

1. Theoretische Kompetenzmodelle

Fragestellung: Wie lassen sich Kompetenzen, unter Berücksichtigung ihres Bezugs auf Anforderungen in spezifischen Situationen, angemessen kognitiv modellieren?

Bei der Formulierung von theoretischen Modellen zur Beschreibung und Erklärung von Kompetenzen geht es zum einen um die Definition von *Strukturen*, zum anderen um die Charakterisierung von *Niveaus*. *Kompetenzstrukturmodelle* befassen sich mit der Frage, welche und wie viele verschiedene Kompetenzdimensionen in einem spezifischen Bereich differenzierbar sind. Bei der Beschreibung von *Kompetenzniveaus* geht es darum, welche konkreten situativen Anforderungen Personen bei welcher Ausprägung einer Kompetenz bewältigen können.

Eine große Herausforderung bei der Formulierung von Kompetenzmodellen besteht darin, dem kontextualisierten Charakter von Kompetenzen gerecht zu werden und sowohl personen- als auch situationsspezifische Komponenten zu berücksichtigen. So können z.B. bei der Beschreibung von Fremdsprachkompetenz die Anforderungen beim Lesen eines Textes und in einem mündlichen Dialog auf Seite der Situation deutlich differenziert werden (z.B. geschriebener vs. gesprochener Text, reines Textverstehen vs. Textproduktion). Auf Seiten der Person müssen zugleich Wissensstrukturen berücksichtigt werden, die in beiden Situationen zum Tragen kommen – etwa das verfügbare Vokabular oder das spezifische Wissen im Bereich grammatischer oder soziopragmatischer Regelungen (Chen, 2004; Kobayashi, 2002). Diese simultane Berücksichtigung personen- und situationsspezifischer Anteile hat sowohl für die Struktur von Kompetenzen als auch für die Beschreibung von Niveaus Konsequenzen. Mit den theoretischen Herausforderungen der Formulierung von Kompetenzmodellen ist schließlich auch die Frage verbunden, wie diese Modelle empirisch angemessen auf ihre Gültigkeit überprüft werden können.

Die Entwicklung theoretisch begründeter und empirisch geprüfter Kompetenzmodelle als Ausgangspunkt für die Entwicklung von psychometrischen Modellen und schließlich von Messverfahren ist z.Zt. immer noch die grundlegende inhaltlich-fachliche Herausforderung der Kompetenzmessung. Es ist erstaunlich, dass dieses Problem bisher noch nicht zufrieden stellend gelöst ist. Auch wenn einzelne (Schul-) Fächer diesbezüglich auf einem guten Weg sind (vgl. z.B. die Arbeiten zum Europäischen Referenzrahmen für Fremdsprachen: Alderson et al., 2005; Alderson, 2005; Gogolin, 2002; die Nutzung von Theorien des mathematischen Modellierens bei PISA: Blum et al. 2004), wird bei der Entwicklung konkreter Aufgabenstellungen zur Diagnostik spezifischer fachbezogener Kompetenzen häufig deutlich, dass der Abstraktionsgrad der bisher

entwickelten Kompetenzmodelle zu hoch ist. Insofern bleibt den Testentwicklern in aller Regel keine andere Wahl, als sehr große Mengen von Aufgaben zu entwickeln, diese Aufgaben empirisch zu erproben und dann im Nachhinein diejenigen Aufgaben beizubehalten, die dem jeweils verwendeten Testmodell entsprechen. Auf diese Weise sind z.B. die bei PISA international berichteten Kompetenzstufenmodelle entstanden (vgl. zum Lesen: Artelt et al., 2001; zur Mathematik: Klieme, Neubrand & Lüdtke, 2001; zu den Naturwissenschaften: Prenzel et al., 2001; zum fächerübergreifendes Problemlösen: Dossey, Hartig, Klieme & Wu, 2004). Die Kompetenzstufen sind allerdings nicht a priori theoretisch begründet, sondern nach Sichtung der testmodell-konform übrig gebliebenen Aufgaben post-hoc definiert worden, was aus theoretischer Perspektive wenig befriedigend ist.

Vor diesem Hintergrund ist in denjenigen Domänen, in denen Kompetenzen diagnostiziert werden sollen, Grundlagenforschung zu theoretisch begründeten und empirisch geprüften Kompetenzstruktur- und Kompetenzentwicklungsmodellen erforderlich. Dies ist primär eine Aufgabe für kognitiv orientierte Fachdidaktiker, Erziehungswissenschaftler und Psychologen. Eine theoriegeleitete Modellierung hat den Vorteil, dass die Verbindung zwischen curricularen Anforderungen, fachspezifischen Zusammenhängen, kognitionspsychologischen Modellen und konkreten Aufgabensammlungen kohärenter gestaltet werden kann. Beispiele finden sich bei Reiss, Hellmich & Thomas (2002) in Bezug auf das Verständnis mathematischer Beweise sowie in mehreren Beiträgen deutscher Autoren zu den TIMSS- und PISA-Skalen, die gerade durch ihren Theoriebezug über das post-hoc-Prinzip der internationalen Berichte hinausweisen. So beschreiben Klieme & Baumert (2001) Leistungen in TIMSS unter Bezug auf Komponenten mathematischen Problemlösens; analog Klieme et al. (2001) sowie Blum et al. (2004) für die beiden PISA-Zyklen. Rost et al. (2003) identifizieren kognitive Komponenten und Repräsentationsmodalitäten im naturwissenschaftlichen Denken (vgl. auch Fischer et al., im Druck), während Artelt et al. (2001) die Textverstehenstheorie von Kintsch und van Dyk zur Modellierung von Lesekompetenz nutzen.

Wichtige Beiträge zur theoriegeleiteten Formulierung von Kompetenzmodellen leistet die Grundlagenforschung aus den Bereichen der kognitiven Psychologie (z.B. Frensch et al., 2003; Haider & Frensch, 1996, 1997, 2002; Hasselhorn & Grube, 2003; Oberauer u.a., 2005; Schneider, Lockl & Fernandez, 2005; Spiel & Glück, im Druck; Weinert & Schneider, 1995), der Neuropsychologie (z.B. Neubauer u.a. 2004; Neubauer u.a., 2005; Nürk, Weger & Willmes, 2005; Stern, im Druck; Willmes, Dambeck & Nürk, 2001) und der differentiellen Psychologie (z.B. Kröner, Plass & Leutner, 2005; Leutner, 2002; Leutner & Plass, 1998; Plass, Chun, Mayer & Leutner, 1998; Wilhelm & Engle, 2005).

Methodische Ansätze für theoriegeleitete Kompetenzmodellierung sind teilweise schon seit längerem bekannt (z.B. das Konzept der Lehrzieltaxonomie, vgl. Leutner, 2001a, in einer neuen Fassung für den kognitiven Bereich: Anderson & Krathwohl, 2001; Verfahren der kognitiven Aufgabenanalyse: Jonassen, Tessmer, & Hannum, 1999; Construct Mapping: Wilson, 2005; Methoden der Novizen-Experten-Forschung: z.B. Boshuizen et al., 2004; Gruber, Renkl & Schneider, 1994). Einen allgemeinen Theorie-

rahmen, mit dem Kompetenzstrukturen, deren Zusammenhang mit beobachtbarem Verhalten in domainspezifischen Situationen und interindividuelle Kompetenzunterschiede dargestellt werden können, bietet auch die so genannte „Kompetenzbasierte Wissensstrukturtheorie“, die während der letzten Dekade von der Arbeitsgruppe um Albert entwickelt wurde (Albert & Held, 1999; Albert & Lukas, 1999).

Ohne Fortschritte bei der theoretischen Darstellung von Kompetenzstrukturen und Kompetenzniveaus werden Diagnostik und Assessment im Bildungsbereich immer wieder auf ein empiristisches Sortieren von Aufgabensammlungen zurückfallen. SPP-Projekte zur Kompetenzmodellierung verfolgen das Ziel, Kompetenzbeschreibungen zu erarbeiten, die (a) der kontextualisierten Natur von Kompetenzen und der Spezifik von Domänen gerecht werden und (b) eine theoriegeleitete Aufgabenentwicklung ermöglichen. Sie können (c) das Verhältnis von kognitiven Prozessmerkmalen (z.B. Lösungsprozessen), Niveaustufung und längsschnittlicher Kompetenzentwicklung bestimmen, und (d) auch nicht-kontinuierliche oder nicht-kumulative Strukturen (z.B. „conceptual change“ beim Erwerb naturwissenschaftlicher Kompetenz durch den Wechsel von Alltags- zu wissenschaftlichen Konzepten; Molz et al., 2004; Schnotz, Vosniadou & Carretero, 1999) zu erfassen suchen. Dies Arbeiten dürften im SPP umso besser gelingen, je stärker kognitionspsychologische Expertise, entweder innerhalb von Projekten oder über Projekte hinweg, mit Domänenexpertise verbunden wird.

2. Psychometrische Modelle

Fragestellung: Wie lassen sich theoretische Kompetenzmodelle in psychometrischen Modellen abbilden, um die Kompetenzkonstrukte einer differenziellen Erfassung zugänglich zu machen?

Aus der theoretischen Modellierung von Kompetenzen ergeben sich Anforderungen an das psychometrische Messmodell – d.h. an das mathematische Regelsystem, das es erlaubt, aus Unterschieden im Testverhalten (z.B. im Muster der ausgewählten Antwortalternativen) auf interindividuelle Unterschiede in der Ausprägung spezifischer Kompetenzen zu schließen (vgl. Hartig, im Druck a; Wilson, 2003). Die bereits auf theoretischer Ebene bestehende Herausforderung, der kontextualisierten Natur von Kompetenzen gerecht zu werden, hat ihre Entsprechung bei der Spezifikation dieser psychometrischen Modelle (vorzugsweise auf Basis der Item-Response-Theorie, IRT): Neben der Modellierung latenter Variablen auf Personenebene müssen auch situationsspezifische Faktoren in die verwendeten Messmodelle einbezogen werden. Eine zentrale Frage ist hierbei, auf welche Weise das Verhältnis zwischen situativen Anforderungen einerseits und Personenmerkmalen andererseits modelliert werden muss, damit individuelle Testwerte als Fähigkeit zur Bewältigung spezifischer Anforderungen interpretiert werden können.

IRT-Modelle zum Einbezug von Aufgabenmerkmalen wie das Linear-Logistische Testmodell (Fischer, 1973) existieren schon länger, wurden bisher jedoch häufiger auf Aufgaben zu kognitiven Grundfähigkeiten angewandt (z.B. Embretson, 1998) als auf

kontextspezifische Aufgaben. Die Untersuchung der Effekte von Aufgabenmerkmalen auf das Antwortverhalten liefert wesentliche Erkenntnisse über die Natur des getesteten Konstrukts und damit über die Validität eines Tests – möglicherweise sogar deutlich wichtigere Informationen als die in psychologischen Validitätsstudien üblichen korrelativen Ansätze (vgl. Embretson, 1998; Borsboom, Mellenbergh & van Heerden, 2004). Eine viel versprechende methodische Entwicklung in diesem Kontext sind „erklärende Item-Response-Modelle“ (*Explanatory Item Response Models*), in denen Personen- und Aufgabenmerkmale simultan Berücksichtigung finden, indem Zufallseffekte auf Itemebene in Messmodelle einbezogen werden (De Boeck & Wilson, 2005; Janssen, Tuerlinckx, Meulders & De Boeck, 2000; Wilson, De Boeck & Carstensen, im Druck). Über Anwendungen dieser Modelle auf Daten aus Kompetenztests wird z.B. von Janssen et al. (2000), Janssen, Schepers und Peres (2004), Hartig und Frey (2005) und Wilson et al. (im Druck) berichtet. Im Zusammenhang mit der DESI-Studie (Beck & Klieme, 2003, im Druck) wurden a priori definierte Aufgabenanforderungen erstmals in das Testmodell eines Large-Scale-Assessments und die Definition der Kompetenzniveaus einbezogen (Hartig, im Druck b).

Bei der psychometrischen Modellierung von Kompetenzen stellt sich weiterhin die Frage, ob interindividuelle Unterschiede angemessener durch quantitative Dimensionen oder durch qualitative Kategorien abgebildet werden können. Im ersteren Fall wären IRT-Modelle mit kontinuierlichen Variablen (z.B. Raschmodell, 2PL, LLTM; vgl. van der Linden & Hambleton, 1996) oder auch mehrdimensionale IRT-Modelle (Ackerman, Gierl & Walker, 2003; Rost & Carstensen, 2002;) ein angemessener Ausgangspunkt, im letzteren Fall IRT-Modelle mit kategorialen latenten Variablen (z.B. Latente Klassenanalyse oder Mischverteilungsmodelle; vgl. Muthén & Muthén, 2000; Rost, 2003; Rost & Langeheine, 1997; Rost & von Davier, 1995). IRT-Modelle mit kategorialen latenten Variablen finden in jüngerer Zeit v.a. in der US-amerikanischen Bildungsforschung zunehmend Verwendung (z.B. von Davier, 2005; von Davier, DiBello & Yamamoto, im Druck; Wilson, 1989; Rupp, 2005). Eine Illustration eines Mischverteilungsmodells zur Abbildung theoretisch postulierter qualitativer Entwicklungssprünge geben z.B. Spiel, Glück & Gößler (2001). Eine weitere Herausforderung besteht schließlich in der Konzeption von Messmodellen, mit denen Kompetenzzuwächse abgebildet und präzise geschätzt werden können (z.B. von Davier, Carstensen & von Davier, im Druck; Hartig & Kühnbach, im Druck; Steyer, 2005).

Die Psychometrie hat – unter international anerkannter Beteiligung deutschsprachiger Wissenschaftler – komplexe Messmodelle entwickelt, die es ermöglichen, kontextualisierte Item- und Personenmerkmale, kategoriale Unterscheidungen und Veränderungen abzubilden. Diese Forschungen sollen in das SPP eingebracht und weitergeführt werden. Das besondere Ziel des SPP im Bereich 2 des Forschungsprogramms (vgl. Abbildung 1) besteht aber – im Einklang mit Pellegrino et al. (2001) – darin, Psychometrie nicht nur als Technik zu nutzen, sondern sie mit der kognitiven Modellierung von Kompetenzen (Bereich 1), der Entwicklung konkreter Messkonzepte und Messverfahren (Bereich 3) und deren Nutzung für Diagnostik und Assessment (Bereich 4) eng zu verzahnen.

3. Messkonzepte und Messverfahren

Fragestellung: *Wie lassen sich Kompetenzmodelle und darauf basierende psychometrische Modelle in konkrete empirische Messverfahren übertragen?*

Die empirische Bildungsforschung, aber auch die unmittelbar an Bildungsprozessen beteiligten Akteure brauchen Messinstrumente, mit denen Kompetenzen empirisch erfassbar gemacht werden können und mit denen handlungsrelevante Rückmeldungen über Ausprägungen und Veränderungen in Kompetenzen gegeben werden können. Zum einen können diese Instrumente in Form standardisierter Testverfahren konstruiert werden, wie sie insbesondere zur Bearbeitung primär wissenschaftlicher Fragestellungen oder in Large-Scale-Assessments eingesetzt werden. Zum anderen kann die Erfassung von Kompetenzen aber auch über systematische Beobachtungen in Bildungsprozessen erfolgen, wie sie z.B. von Lehrkräften an Schülern vorgenommen werden. Auch im Hinblick auf diese Methoden besteht Forschungsbedarf; Stichwörter sind in diesem Zusammenhang z.B. Portfolio-Assessment (Bennett & Ward, 1993; Lissmann, 2004) oder die Vergleichbarkeit von Schulabschlüssen und Schulnoten (Köller, Baumert & Schnabel, 1999; Köller, Watermann, Trautwein & Lüdtke, 2004).

Die Pädagogische Diagnostik als das Arbeitsfeld, das sich traditionell mit dem Einsatz von Messverfahren (insbesondere Tests) zur empirischen Erfassung von Kompetenzen befasst, hatte international wie auch in Deutschland in den 1970er-Jahren eine besonders intensive Entwicklungsphase (Klauer, 1978). Im Kontext von Bildungsreformen, insbesondere Curriculumrevisionen mit der Einführung neuer Lehr- bzw. Lernziele sowie der Etablierung und Evaluation neuer Bildungsgänge (integrierte Gesamtschulen, Gymnasiale Oberstufe) wurden, entsprechend dem Konzept der lehrziel- bzw. kriteriumsorientierten Leistungsmessung (vgl. Klauer, 1987), Testaufgaben systematisch entwickelt und Testskalen gebildet, die in der individuellen Förderdiagnostik, aber auch zur Unterrichts- und Schulevaluation eingesetzt wurden. Zugleich wurden neue, probabilistische Testtheorien entwickelt und erprobt, mit denen der Zusammenhang zwischen beobachtetem Antwortverhalten und latenter Fähigkeit modelliert werden konnte. Diese Forschungstradition ist in anderen Ländern, wie z.B. den Niederlanden und in den USA, auf breiter Ebene kontinuierlich fortgeführt worden (z.B. van der Linden & Hambleton, 1996).

In Deutschland kam die Pädagogische Diagnostik, insbesondere im Hinblick auf die praktische Anwendung neuerer Entwicklungen, nach den 1970er-Jahren weitgehend zum Erliegen (vgl. Klauer, 1994). Dies betrifft z.B. das computerbasierte adaptive Testen, zu dem zuvor wichtige Beiträge geleistet worden waren (z.B. Hornke, 1999; Kubinger, 1987). Erst in den 1990er-Jahren fand die deutschsprachige Pädagogische Diagnostik durch die verantwortliche Mitgestaltung von Testverfahren für internationale Schulleistungsvergleichsstudien wieder Anschluss an den internationalen „state of the art“. So waren bei PISA 2000 deutsche Wissenschaftlerinnen und Wissenschaftler aktiv in *internationale* Expertengruppen eingebunden und an der Konzeptionsentwicklung beteiligt, insbesondere im Bereich der Messung von Selbstregulationsfähigkeit (Artelt, Baumert, Julius-McElvany & Peschar, 2003). In PISA 2003 wurde mit dem Bereich des „cross-

curricular problem solving“ (Dossey et al., 2004; Klieme et al., 2005; Leutner et al., 2004, 2005; Leutner & Wirth, im Druck) eine Testkomponente international eingesetzt, die zuvor in Deutschland entwickelt und sowohl international im Rahmen des Adult Literacy and Lifeskills Survey (ALL; Reeß, Zabal & Klieme, 2004) als auch als nationale Ergänzung zu PISA 2000 (Frensch & Funke, 1995; Funke, 2003; Klieme et al., 2001, 2005; Wirth & Klieme, 2003) erprobt worden war. Bei PISA 2006 schließlich ist mit dem IPN in Kiel ein deutsches Institut im internationalen Konsortium für den Schwerpunkt der Testentwicklung (hier: Naturwissenschaften) zuständig. Für PISA 2009 wird aus dem Kreis der zu erwartenden Teilnehmer und Initiatoren des SPP ebenso eine Beteiligung deutscher Institute an der Testentwicklung vorbereitet wie für das geplante OECD-“Programme for International Adult Assessment of Competencies“ (PIAAC). Aktuelle innovative Arbeiten zur Konstruktion modellbasierter Messinstrumente wurden auch in der DESI-Studie geleistet (Beck & Klieme, 2003, im Druck), z.B. für Englisch als Fremdsprache (vgl. Nold, 2003; Nold & Rossa, im Druck a, b). Das SPP soll dazu beitragen, diese international besonders sichtbaren Beiträge deutscher Bildungsforschung im Hinblick auf thematisch einschlägige Grundlagenforschung zu bündeln, sie auf diese Weise zu stärken und damit international noch sichtbarer zu machen. Besonderes Augenmerk wird dabei auch auf innovativen Messverfahren liegen, die sich in zunehmendem Maße neuer Technologien bedienen – sei es durch komplexe Aufgabenformate, durch interaktive Testdurchführung und in Echtzeit erhobene Prozessmaße (Wirth, 2004; Wirth & Klieme, 2003) oder durch automatisierte Auswertungs- und Rückmeldeprozeduren (Chung, O’Neil & Baker, im Druck; Ordinate, 2004; Reeß & Martin, im Druck). Computer- und internetbasierte Messung bieten darüber hinaus eine besondere Chance für die präzise Individualdiagnostik (adaptives Testen, z.B. van der Linden, 2005). Die genannten Vorteile kommen in kompetenzbezogenen Erhebungen besonders zur Geltung, weil sie es erlauben, Leistungen zu kontextualisieren, z.B. durch den Einsatz von Computersimulationen komplexer Situationen.

Die Konstruktion von Messkonzepten und Messverfahren in verschiedenen Kompetenzdomänen ist ein wichtiges Ziel des SPP. Eine erfolgreiche Bearbeitung wird u.a. auch die Chancen erhöhen, dass deutsche Wissenschaftler eine führende Rolle bei zukünftigen internationalen Leistungsstudien wie PISA 2009 und PIAAC einnehmen. Die Arbeit in diesem Bereich ist innerhalb des SPP jedoch kein Selbstzweck. Vielmehr soll exemplarisch gezeigt werden, dass sich innovative Kompetenzmodelle und auf deren Basis entwickelte psychometrische Modelle in konkrete Messverfahren übertragen lassen, welche sowohl als Erhebungsinstrumente in weiterführender Grundlagenforschung als auch in angewandten Kontexten eingesetzt werden können. Im Vordergrund stehen nicht die technischen Details, sondern neue Messkonzepte.

4. Nutzung von Informationen aus Diagnostik und Assessment

Fragestellung: Welche Arten von Informationen aus Kompetenzmessungen können von Akteuren im Bildungswesen auf welche Weise genutzt werden?

Kompetenzdiagnostik erfolgt im Dienste personenbezogener Entscheidungen; Assessment-Programme in Bildungsinstitutionen hingegen zielen auf die Qualitätsentwicklung in diesen Einrichtungen. Aus der Spannung zwischen diesen Nutzenperspektiven ergeben sich vor allem erziehungswissenschaftlich relevante Forschungsfragen: Obwohl in der gegenwärtigen öffentlichen Diskussion Bildungsstandards und Bildungs-Monitoring-Aspekte des Large-Scale-Assessments dominieren (z.B. Helmke & Hosenfeld, 2004; Stanat & Baumert, 2001), ist zu fragen, wie diese eher bildungspolitischen und administrativen Interessen mit den eher pädagogisch bzw. didaktisch orientierten Interessen der Schulen und Lehrpersonen und ihrer Klassen vor Ort in Einklang gebracht werden können. Dies betrifft insbesondere auch Fragen nach der Legitimation weit reichender Konsequenzen kompetenzdiagnostischer Maßnahmen (im Englischen: „high-stakes assessment“) für einzelne Schülerinnen und Schüler (z.B. Bildungsabschlüsse), Lehrpersonen (z.B. Beförderungen) oder Schulen (z.B. Erhalt vs. Schließung). In der internationalen Forschungsliteratur liegen Konzepte zur Verknüpfung von Diagnostik und Assessment auf unterschiedlichen Ebenen (Scheerens, Glas & Thomas, 2003) sowie – noch uneinheitliche – empirische Befunde zu den Auswirkungen von Assessment-Systemen vor (Cheng et al., 2004; Fuhrman & Elmore, 2004). Völlig offen ist, wie solche Systeme sich im deutschen Schulwesen mit seinen spezifischen Ansprüchen an professionelle Autonomie auswirken. Hier öffnet sich ebenfalls ein weites Feld offener Forschungsfragen: Wer diagnostiziert auf der Basis welcher diagnostischen Kompetenz? Wer trifft Entscheidungen auf der Basis welcher Entscheidungsstrategien? Welche Auswirkungen hat beispielsweise die Teilnahme an Schulleistungsvergleichen oder/und zentralen Abschlussprüfungen auf die diagnostische Praxis und Kompetenz von Lehrpersonen? Inwieweit können Messergebnisse aus einem Large-Scale-Assessment für individuelle Förderentscheidungen herangezogen werden? Möglicherweise bestehen für eine kompetenzbezogene Diagnostik besondere Chancen und Risiken: Welche Vorstellungen haben Lehrende – implizit oder explizit – von Kompetenzmodellen? Analoge Fragen könnten sich für andere Bildungsbereiche (z.B. Hochschule und berufliche Bildung) stellen, u.a. aufgrund zunehmender Verbreitung von Zulassungsverfahren und modularen Ausbildungsstrukturen.

Projekte zu Fragen der Nutzung von Diagnostik und Assessment arrondieren das SPP mit Blick auf pädagogische Handlungsfelder und bildungspolitische Erwägungen. Sie stehen zwar nicht im Zentrum des SPP, sind aber unverzichtbar, um die Kompetenzmessung selbst zu situieren. Entsprechende Projekte werden prüfen, unter welchen Nutzungsbedingungen modellbasierte Kompetenzmessung auch auf der angewandten Seite einen Mehrwert an Präzision und Validität pädagogischer Entscheidungen erzeugt. Beispielsweise sind Projekte willkommen, die sich mit Effekten verschiedenartiger Messmethoden (z.B. Kompetenzstufenmodelle vs. traditionelle, normorientiert konstruierte Schulleistungstests) in einem Praxisfeld bzw. einem Messverfahren in verschiedenen Praxisfeldern befassen. Eine entscheidende Größe in solchen Untersuchungen wird die diagnostische Kompetenz der im jeweiligen Praxisfeld Lehrenden sein.

Vorarbeiten

Zur Vorbereitung des Schwerpunktprogramms wurde in einem von J. Hartig und den Initiatoren herausgegebenen Sammelband über „Assessment of Competencies in Educational Contexts: State of the Art and Future Prospects“ (Hartig, Klieme & Leutner, im Druck) eine Aufarbeitung des Forschungsstandes und eine Sichtung offener Fragestellungen vorgenommen. Weiterhin wurden über die einschlägigen wissenschaftlichen Organisationen empirisch arbeitende Erziehungswissenschaftler und Fachdidaktiker sowie an Bildungsprozessen interessierte Psychologen aufgefordert, Projektskizzen zu entwerfen. Zahlreiche (ca. 300), zu einem großen Teil DFG-erfahrene und international ausgewiesene Kolleginnen und Kollegen, darunter auch sehr viele Nachwuchswissenschaftler, haben reagiert und deutlich gemacht, dass sie vor dem Hintergrund eigener Vorarbeiten in der Lage sind, kompetenzdiagnostische Forschungsvorhaben auf theoretisch und methodisch hohem Niveau zu konzipieren. In ihrer Gesamtheit decken die ca. 150 Projektskizzen das in Frage stehende Forschungsprogramm hervorragend ab. Für die Implementation des Schwerpunktprogramms bedeutet dies, dass mit einer großen Anzahl an ausgearbeiteten Projektanträgen gerechnet und somit sichergestellt werden kann, dass an Exzellenzkriterien orientierte Forschung in der für ein SPP erforderlichen Breite tatsächlich auf den Weg zu bringen ist. Eine Präzisierung des Forschungsprogramms wurde schließlich im Oktober 2005 im Rahmen eines Rundgesprächs bei der DFG mit Repräsentanten der einschlägigen Fachdisziplinen vorgenommen.

Zu den Vorarbeiten gehört darüber hinaus, dass in enger Kooperation der Initiatoren mit dem BMBF konkrete Planungen bestehen, das DFG-SPP mit einem parallel einzurichtenden Forschungs- und Entwicklungsprogramm des BMBF zum Thema „Technologie-gestütztes Assessment“ zu koordinieren. Das technologisch orientierte BMBF-Programm soll durch die Ergebnisse der DFG-geförderten Grundlagenforschung untermauert und auf eine wissenschaftlich solide Basis gestellt werden; andererseits wird angestrebt, die im BMBF-Programm erfolgten technologischen Entwicklungen in den Projekten des SPP zu nutzen, um z.B. Kompetenzdaten zu erheben und für die weitere statistische Analyse aufzubereiten. Eine für diese Zwecke geeignete internet-basierte Plattform zur Organisation, Vorgabe und Auswertung von Testaufgaben wurde an der Universität Luxemburg entwickelt und erfolgreich erprobt (Martin, Latour, Burton, Busana & Vandenabeele, 2005; Plichart u.a., im Druck).

Zu erwartende Teilnehmer und Projekte

Den Initiatoren sind bei der Vorbereitung des SPP bereits zahlreiche Projektskizzen eingereicht worden. Nach der öffentlichen Ausschreibung des SPP durch die DFG werden weitere Skizzen folgen. Auf der Basis ausgearbeiteter Projektanträge sollen Kolleginnen und Kollegen dann in das SPP aufgenommen werden, die über einschlägig veröffentlichte Vorarbeiten verfügen und deren Projektvorschläge in den engeren thematischen Rahmen des SPP hineinpassen. Besonders angesprochen fühlen sollten sich führende

Vertreter der für die Arbeiten im SPP erforderlichen Fachdisziplinen aus der Psychologie, der Erziehungswissenschaft und den Fachdidaktiken, die bereits über DFG-Erfahrung verfügen und mit ihrer Forschung – je nach den Standards ihrer Disziplin – sowohl national als auch international sichtbar sind.

Unter den zu erwartenden Teilnehmern des SPP wird es zahlreiche Netzwerke aus interdisziplinär erfolgreichen Kooperationen in Konsortien und Expertengruppen von Large-Scale Assessments geben (international: TIMSS, PISA und DESI; national: Vergleichsarbeiten und Lernstandserhebungen in den Bundesländern). Darüber hinaus werden viele der zu erwartenden Teilnehmer im Zusammenhang mit der Definition und Überprüfung nationaler Bildungsstandards in Deutschland kooperieren. Diese vielfältig verzahnten, größtenteils auch interdisziplinären Kontakte auf der Arbeitsebene sind nicht nur eine solide Basis für Kooperation und Netzwerkbildung im SPP, sondern darüber hinaus auch eine Gewähr dafür, dass die Forschungsergebnisse des SPP neben theoretischen auch unmittelbar praktische Implikationen haben werden.

Gestaltung des Programms

Interdisziplinäre Zusammenarbeit

Das SPP soll die für eine an wissenschaftlichen Qualitätsstandards orientierte Kompetenzmodellierung erforderlichen Disziplinen zusammenbringen: (1) Fachdidaktiker als Experten für bereichsspezifische Kompetenzen und Bildungsprozesse, (2) pädagogische Psychologen und Erziehungswissenschaftler als Experten für bereichsübergreifende Aspekte von Kompetenzen, Bildungsprozessen und deren Voraussetzungen sowie (3) Psychometriker als Experten für Diagnostik und Assessment. Auch auf internationaler Ebene beginnt sich mittlerweile die Sichtweise durchzusetzen, dass keine der beteiligten Disziplinen allein dazu imstande wäre, die anstehenden Fragestellungen angemessen zu bearbeiten. Insofern erfordert das SPP interdisziplinäre Zusammenarbeit und – weil die Disziplinen nicht an allen potentiellen Standorten auf dem erforderlichen wissenschaftlichen Niveau vertreten sind – ortsübergreifende Zusammenarbeit und Netzwerkbildung. Zusätzlich zu bereits bestehenden Kooperationen zwischen den zu erwartenden Teilnehmern wird angestrebt, thematische Netzwerke innerhalb des SPP zu bilden. Entsprechend der inhaltlichen Fragestellungen des SPP sollen sich Netzwerke bilden, die sich mit verwandten Kompetenzkonzepten oder mit ähnlichen methodischen Fragestellungen befassen. Dadurch werden sich u.a. auch sehr spezifische Aspekte von Mehrwert ergeben, z.B. im Hinblick auf einen theoretisch fundierten Abgleich von Kompetenzmodellen in verschiedenen Kompetenzdomänen.

Im Rahmen einer Vorphase sollen die Antragsteller zum Zeitpunkt der Antragstellung angehalten werden, Vernetzungsmöglichkeiten auszuloten. Dies ist vor Aufnahme der Arbeit zu konsolidieren, wenn bekannt ist, welche Projekte in das SPP aufgenommen worden sind. Möglicherweise lassen sich dann zusätzliche Synergien und thematische Verknüpfungen schaffen, indem empirische Erhebungen gemeinsam genutzt wer-

den. Regelmäßige Rundgespräche (zwei je Jahr) sollen den Austausch zwischen allen Teilnehmern des SPP forcieren. Gezielter Austausch von Wissenschaftlern („Lab-Rotation“) soll die interdisziplinäre Zusammenarbeit projektübergreifend fördern. Möglichkeiten für die Einrichtung von Methodenprojekten „quer“ zu stärker auf Inhaltsdomänen fokussierten Einzelprojekten im Sinne einer Zusammenarbeit von Inhaltsexperten und Psychometrikern über Projektgrenzen hinweg sind zum Zeitpunkt der Antragstellung auszuloten.

Internationale Vernetzung und Sichtbarkeit

Die Initiatoren und andere an der Vorbereitung des SPP beteiligten Kolleginnen und Kollegen verfügen über vielfältige, insbesondere auch durch Veröffentlichungen dokumentierte Kontakte mit Kollegen und Forschungsinstitutionen im europäischen und außereuropäischen Ausland. Hierzu haben auch die Beteiligungen deutscher Wissenschaftler an internationalen Assessment-Programmen (v.a. PISA) beigetragen (vgl. Abschnitt „Messkonzepte und Messverfahren“). Hervorzuheben sind insbesondere kontinuierliche Arbeitskontakte mit Zentren der Kompetenzmessung wie ACER (Melbourne; Zentrale der PISA-Studien), CITO (Arnheim; wichtigste niederländische Testagentur), CRESST (Center for Research on Evaluation, Standards and Student Testing, UCLA, Los Angeles), ETS (Educational Testing Service, Princeton), und dem Berkeley Evaluation & Assessment Research Center (BEAR; Prof. Mark Wilson), die von Seiten des DIPF und des IPN, aber auch des IQB (Köller, HU Berlin) und anderer Beteiligter bestehen. Diese Kontakte haben ihren Niederschlag u.a. in Beiträgen von Experten aus diesen Institutionen für den vorbereitenden Sammelband (Hartig, Klieme & Leutner, im Druck) gefunden.

In einigen bei der Vorbereitung des SPP diskutierten Projektvorschlägen sind bereits internationale Kooperationen mit inhaltlich und methodisch hervorragend ausgewiesenen Kollegen geplant. Die bestehenden internationalen Kontakte der Initiatoren und zu erwartenden SPP-Teilnehmer sollen zudem genutzt werden, international renommierte Kollegen als „Critical Friends“ zu gewinnen, die dem SPP in Form eines wissenschaftlichen Beirats zur Seite stehen. Dieser Beirat soll einmal jährlich an einem Rundgespräch teilnehmen, zu den Projektdarstellungen und Forschungsvorhaben der Nachwuchswissenschaftler Stellung nehmen und sich ggf. mit Vorträgen beteiligen.

Konzepte zur Einbindung und Förderung des wissenschaftlichen Nachwuchses

Die Förderung des wissenschaftlichen Nachwuchses soll eine zentrale Rolle im SPP spielen. Unter anderem soll dies bereits zu Beginn dadurch realisiert werden, dass promovierte junge Kolleginnen und Kollegen ermutigt werden, eigene Projektanträge in das Schwerpunktprogramm einzubringen. In diesem Zusammenhang soll das SPP einige jüngere deutsche Wissenschaftler, die im Ausland tätig sind, als Teilnehmer einbinden und für eine wissenschaftliche Tätigkeit in Deutschland zurück zu gewinnen suchen.

Im SPP soll ein regelmäßiger Austausch zwischen den Nachwuchswissenschaftlern (insbesondere den Doktoranden) gewährleistet werden. Ein Instrument sind Doktoranden-Meetings bei den regelmäßigen Rundgesprächen des SPP, auf denen diese ihre Forschung vorstellen und diskutieren. Ein weiteres Instrument sind wechselseitige Hospitationen von Doktoranden verschiedener Projekte, um Einblicke in die Arbeitsweisen anderer Fachdisziplinen zu erhalten.

Die Doktoranden sollen ihre Dissertationsvorhaben wenigstens einmal bei einem SPP-Rundgespräch im Plenum aller Teilnehmer vorstellen. Zur Weiterbildung sollen Workshops organisiert werden, in denen die in den beteiligten Fachdisziplinen verwendeten wissenschaftlichen Ansätze und Forschungsmethoden vermittelt werden. Für diese Workshops sollen zum einen externe Referenten gewonnen werden, zum anderen sollen die Nachwuchswissenschaftler aber auch selbst ermutigt werden, ihre eigenen Kenntnisse in diesen Workshops an Kollegen zu vermitteln. Darüber hinaus werden die Doktoranden angehalten, an den im Rahmen des DFG-Förderprogramms „Empirische Bildungsforschung“ jährlich geplanten interdisziplinären Sommerschulen teilzunehmen, um insbesondere auch ihre forschungsmethodischen Kenntnisse zu komplettieren. Dabei sollen Kontakte zu international renommierte Kollegen genutzt werden, um den wissenschaftlichen Nachwuchs im SPP hinsichtlich Forschungs- und Publikationsstrategien zu beraten.

Zur Förderung des wissenschaftlichen Nachwuchses werden sowohl die angestrebten Netzwerke innerhalb des Schwerpunkts als auch ausländische Wissenschaftler und Forschungsinstitutionen eine wichtige Rolle spielen. Durch die Einbindung internationaler Experten als „Critical Friends“ in das Schwerpunktprogramm sollen sowohl für Doktoranden als auch für Post-Doc-Wissenschaftler Praktika oder Gastaufenthalte an internationalen, für die Bildungsforschung und Kompetenzdiagnostik relevanten Institutionen (z.B. ETS, CRESST, ACER, CITO) ermöglicht werden.

Abgrenzung zu anderen Förderaktivitäten der DFG und sonstigen Forschungsverbünden

Viele der zu erwartenden SPP-Teilnehmer werden an Forschungs- und Entwicklungsprojekten im Bereich der Kompetenzdiagnostik beteiligt sein (s.o.). Die dahinter liegenden Forschungsverbünde (international: z.B. TIMSS, PISA, IGLU, DESI; national: Qualitätsagenturen der dt. Bundesländer und Kooperationen mit dem BMBF) betreiben primär angewandte Forschung, werden aber von den grundlagenwissenschaftlich orientierten Ergebnissen des geplanten SPP erheblich profitieren können. Darüber hinaus gibt es einige Förderaktivitäten der DFG, von denen sich das geplante SPP deutlich abgrenzen lässt:

Schwerpunkt 1043. Lesesozialisation in der Mediengesellschaft: Geschlechtsspezifische/-übergreifende Strukturen, Prozesse, Bedingungsbeziehungen: Befasst sich u.a. mit Lesen als Schlüsselkompetenz, aber nicht mit der Diagnostik derselben und nur in wenigen Projekten mit Bildungsprozessen. Endet 2008.

Schwerpunkt 1081. Netzbasierte Wissenskommunikation in Gruppen: Befasst sich nicht wesentlich mit Assessment. Endet 2006.

Schwerpunkt 1082. Die Bildungsqualität von Schule: Fachliches und fächerübergreifendes Lernen im mathematisch-naturwissenschaftlichen Unterricht in Abhängigkeit von schulischen und außerschulischen Kontexten: Dort sind wesentliche Vorarbeiten geleistet worden, auf die das hier beantragte SPP aufbauen kann. Ein Schwerpunkt von „Bildungsqualität von Schule“ lag auf den Vermittlungsbedingungen von Kompetenzen im schulischen Kontext, aber nicht auf deren Messung. Endet 2006.

Forscherguppe 454. Heterogene Arbeit: Positive und normative Aspekte der Qualifikationsstruktur: Befasst sich mit Qualifikationen aus Sicht des Arbeitsmarktes, nicht mit der Definition und Erfassung auf individueller Ebene.

Forscherguppe 543. Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor- und Grundschulalter (BiKS): Der Fokus ist enger gefasst und liegt nicht auf der Definition und Erfassung von Kompetenzen, sondern auf deren Entwicklung in einem spezifischen Altersabschnitt.

Antragszeitraum, geplante Förderperioden und Mittelbedarf

Das SPP soll 2007 seine Arbeit aufnehmen. Geplant sind drei Förderperioden á zwei Jahre, um die Einhaltung von DFG-Qualitätsstandards salient zu machen. Die Projekte sollen DFG-übliche Größenordnungen mit einem (bei einigen Projekten, sofern das geplante Arbeitsprogramm es rechtfertigt, auch zwei oder mehr) wissenschaftlichen Mitarbeitern auf Qualifizierungsstellen (BAT IIa/2) haben. Hinzu kommen studentische Hilfskräfte im Umfang von 10-20 Stunden/Woche je wissenschaftlichem Mitarbeiter, Sachmittel, Reisekosten zum Austausch zwischen den Einzelprojekten und zu Rundgesprächen sowie Reisekosten zu nationalen und internationalen Kongressen. Zu berücksichtigen ist, dass für die empirische Untersuchung von Kompetenzen ggf. spezifische Stichproben rekrutiert werden müssen (z.B. Schüler oder Berufstätige in bestimmten Tätigkeitsfeldern), die für die sinnvolle Anwendung moderner Analyseverfahren (z.B. IRT- oder Mehrebenenmodelle) gewisse Mindestgrößen haben müssen, was besondere Kosten der Datenerhebung und Datenauswertung verursachen kann. Zusätzlich zu den Einzelprojekten wird ein Koordinationsprojekt erforderlich sein.

Literatur

- Ackerman, T. A., Gierl, M. J. & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, 22, 37-53.
- Albert, D. & Held, T. (1999). Component-based knowledge spaces in problem solving and inductive reasoning. In A. Dietrich & L. Josef (Eds.), *Knowledge spaces: Theories, empirical research, and applications* (pp. 15-40). Mahwah: Lawrence Erlbaum Associates.
- Albert, D. & Lukas, J. (Eds.) (1999). *Knowledge spaces: Theories, empirical research, and applications*. Mahwah: Lawrence Erlbaum Associates.

- Alderson, C. (2005). Diagnosing foreign language proficiency: the interface between learning and assessment. London: Continuum.
- Alderson, C., Figueras, N., Kuijper, H., Nold, G., Takala, S. & Tardieu, C. (2005). The Dutch CEF rid Reading / Listening (Revised internet version available for test development and analysis). <http://www.lancs.ac.uk/fss/projects/grid/>.
- Anderson, L. W. & Krathwohl, D. R. (2001). A taxonomy for learning, teaching and assessing. A revision of Bloom's taxonomy of educational objectives. New York: Addison Wesley Longman, Inc.
- Artelt, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). Learners for life. Student approaches to learning. Results from PISA 2000. Paris: OECD Publications.
- Artelt, C., Schiefele, U. & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education*, 16, 363-383.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In: Deutsches PISA-Konsortium (Hrsg.), PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich (S. 69.-137). Opladen: Leske und Budrich.
- Baumert, J., Evans, R. H., Geiser, H. (1998). Technical problem solving among 10-year-old students as related to science achievement, out-of-school experience, domain-specific control beliefs, and attribution patterns. *Journal of Research in Science Teaching*, 35, 987-1013.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Hrsg.), PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich (S. 15-68). Opladen: Leske & Budrich.
- Beck, B. & Klieme, E. (2003). DESI – Eine Large-Scale-Studie zur Untersuchung des Sprachunterrichts in deutschen Schulen. *Zeitschrift für empirische Pädagogik*, 17, 380-395.
- Beck, B. & Klieme, E. (im Druck). Sprachliche Kompetenzen. Konzepte und Messung in der DESI-Studie. Weinheim: Beltz.
- Bennett, R. E. & Ward, W. (Eds.) (1993). Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment. Hillsdale: Lawrence Erlbaum Associates.
- Blömeke, S. (2003). Medienpsychologische Kompetenz. Theoretische Grundlagen und erste empirische Befunde. *Empirische Pädagogik*, 17, 196-216.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. & Carstensen, H. C. (2004). Mathematische Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs (S. 47-92). Münster: Waxmann.
- De Boeck, P. & Wilson, M. (2004). Explanatory item response models. A generalized linear and nonlinear approach. Berlin: Springer.
- Borsboom, D., Mellenbergh, G. J. & van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111, 1061-1071.
- Boshuizen, H. P. A., Bromme, R. & Gruber, H. (Eds.). (2004). Professional learning: Gaps and transitions on the way from novice to expert. Dordrecht: Kluwer.
- Breuer, K. & Satish, U. (2003). Emergency Management Simulations – An approach to the assessment of decision making processes in complex dynamic crisis environments. In González, José J. (Ed.), From modeling to managing security – A system dynamics approach (pp. 145-156). Kristiansand: Norwegian Academic Press.
- Bruder, R., Perels, F., Gürtler, T. & Schmitz, B. (2002). Trainingsprogramm zur Förderung sachspezifischer und fächerübergreifender Kompetenzen in der achten Jahrgangsstufe von Gymnasien. *Mathematik Lehren*, 115, 59-62.
- Chen, L. (2004). On text structure, language proficiency, and reading comprehension test format interactions: a reply to Kobayashi, 2002. *Language Testing* 21, 228-234.

- Cheng, L., Watanabe, Y. & Curtis, A. (Eds.). (2004). Washback in language testing. research contexts and methods. Mahwah: Lawrence Erlbaum.
- Chung, G. K. W. K., O'Neil, H. F. & Baker, E. L. (im Druck). Computer-based assessments to support distance learning. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts: state of the art and future prospects*. Göttingen: Hogrefe & Huber.
- Connell, M. W., Sheridan, K. & Gardner, H. (2003). On abilities and domains. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 126-155). Cambridge: Cambridge University Press.
- Csapó, B. (2004). Knowledge and competencies. In J. Letschert (Ed.), *The integrated person. How curriculum development relates to new competencies* (pp. 35-49). Enschede: CIDREE/SLO.
- von Davier, M. (2005). A general diagnostic model applied to language testing data. ETS Research Report 0x-2005.
- von Davier, A., Carstensen, C. H. & von Davier, M. (im Druck). Linking competencies in horizontal, vertical and longitudinal settings and measuring growth. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts: state of the art and future prospects*. Göttingen: Hogrefe & Huber.
- von Davier, M., DiBello, L. & Yamamoto, K. (im Druck). Feedback of test outcomes using models for cognitive diagnosis. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts: state of the art and future prospects*. Göttingen: Hogrefe & Huber.
- Ditton, H. (2000): Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. In A. Helmke, W. Hornstein & E. Terhart (Hrsg.), *Qualität und Qualitätssicherung im Bildungsbereich. Zeitschrift für Pädagogik*, 41. Beiheft, 73-92.
- Dossey, J., Hartig, J., Klieme, E. & Wu, M. (2004). Problem solving for tomorrow's world. First measures of cross-curricular competencies from PISA 2003. Paris: OECD Publications.
- Dutke, S. (2000). Multiple representations and individual differences in generating mental models: The case of text comprehension. In U. von Hecker, S. Dutke & G. Sedek (Eds.), *Generative mental processes and cognitive resources: Integrative research on adaptation and control* (pp. 67-93). Dordrecht, The Netherlands: Kluwer.
- Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: application to abstract reasoning. *Psychological Methods*, 3, 380-396.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.
- Fischer, H. E., Klemm, K., Leutner, D., Sumfleth, E., Thiemann, R. & Wirth, J. (im Druck). Framework for empirical research on science teaching and learning. *Journal of Science Teacher Education*.
- Frensch, P. A. & Funke, J. (1995). Complex problem solving. The European perspective. Hillsdale, NJ: Erlbaum.
- Frensch, P. A., Haider, H., Rüniger, D., Neugebauer, U., Voigt, S. & Werg, D. (2003). The route from implicit learning to awareness of what has been learned. In L. Jiménez (Ed.), *Attention and implicit learning*. New York: John Benjamins Publishing Company.
- Fuhrman, S. H. & Elmore, R. F. (Eds.) (2004). *Redesigning accountability systems for education*. New York: Teachers College Press.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Gogolin, I. (2002). Linguistic and cultural diversity in Europe: a challenge for educational research and practice. *European Educational Research Journal*, 1, 123-138.
- Gold, A. & Souvignier, E. (in Druck). Prognose der Studierfähigkeit. Ergebnisse aus Längsschnittanalysen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*.

- Grube, D. & Hasselhorn, M. (in Druck). Längsschnittliche Analysen zur Entwicklung des Lesens, Rechtschreibens und Rechnens im Grundschulalter. In I. Hosenfeld & F.-W. Schrader (Hrsg.), *Schulische Leistung: Grundlagen, Bedingungen, Perspektiven*. Münster: Waxmann.
- Gruber, H. (2004). Kompetenzen von Lehrerinnen und Lehrern – Ein Blick aus der Expertiseforschung. In A. Hartinger & M. Fölling-Albers (Hrsg.), *Lehrerkompetenzen für den Sachunterricht* (S. 21-33). Bad Heilbrunn: Klinkhardt.
- Gruber, H., Harteis, C., Heid, H. & Meier, B. (Hrsg.) (2004). *Kapital und Kompetenz. Veränderungen der Arbeitswelt und ihre Auswirkungen aus erziehungswissenschaftlicher Sicht*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gruber, H., Renkl, A. & Schneider, W. (1994). Expertise and memory development: Longitudinal comparisons in the field of chess. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 26, 53-70.
- Haider, H. & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, 30, 304-337.
- Haider, H. & Frensch, P. A. (1997). Lernmechanismen des kognitiven Fertigkeitserwerbs. *Zeitschrift für Experimentelle Psychologie*, 44, 521-560.
- Haider, H., & Frensch, P. A. (2002). Why individual learning does not follow the power law of practice but aggregated learning does: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 392-406.
- Hartig, J. (im Druck a). Latent variable modelling in educational assessment. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts: State of the Art and Future Prospects*. Göttingen: Hogrefe & Huber.
- Hartig, J. (im Druck b). Skalierung und Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung in der DESI-Studie*. Weinheim: Beltz.
- Hartig, J. & Frey, A. (2005). Application of different explanatory item response models for model based proficiency scaling. Paper presented at the 70th Annual Meeting of the Psychometric Society in Tilburg, July 5–8, 2005.
- Hartig, J. & Klieme, E. (im Druck). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik*. Berlin: Springer.
- Hartig, J., Klieme, E. & Leutner, D. (Eds.) (im Druck). *Assessment of competencies in educational contexts: state of the art and future prospects*. Göttingen: Hogrefe & Huber.
- Hartig, J. & Kühnbach, O. (im Druck). Schätzung von Veränderung mit Plausible Values in mehrdimensionalen Rasch-Modellen. In A. Ittel & H. Merckens (Hrsg.), *Veränderungsmessung und Längsschnittstudien in der empirischen Erziehungswissenschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hartinger & Fölling-Albers, M. (Hrsg.). (2004). *Lehrerkompetenzen für den Sachunterricht. Probleme und Perspektiven des Sachunterrichts Bd. 14* (S.9-18). Bad Heilbrunn: Klinkhardt.
- Hasselhorn, M. & Grube, D. (2003). The phonological similarity effect on memory span in children: Does it depend on age, speech rate, and articulatory suppression? *International Journal of Behavioral Development*, 27, 145-152.
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, 54, 3-56.
- Helmke, A. & Hosenfeld, I. (2004). Vergleichsarbeiten – Kompetenzmodelle – Standards. In M. Wosnitza, A. Frey & R. S. Jäger (Hrsg.), *Lernprozesse, Lernumgebungen und Lerndiagnostik. Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert* (S. 56-75). Landau: Verlag Empirische Pädagogik.
- Herzmann, P., Sparka, A. & Gräsel, C. (im Druck). Implementationsforschung zur Lesekompetenz: Wie Wissenschaftler und Lehrkräfte gemeinsam an der Lehrerförderung arbeiten. In S. Rahm (Hrsg.), *Innovative Ansätze der Schul- und Unterrichtsforschung*.

- Holling, H., Vock, M., & Wittmann, A. J. (2001). Wie erkennt man Hochbegabung bei jungen Erwachsenen? In Deutsche Gesellschaft für das hochbegabte Kind e.V. (Hrsg.), *Im Labyrinth: Hochbegabte Kinder in Schule und Gesellschaft* (S. 16-21). Münster: LIT.
- Hornke, L. (1999). Benefits from computerized adaptive testing as seen in simulation studies. *European Journal of Psychological Assessment*, 15, 91-98.
- Janßen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In P. De Boeck, & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 189-212). New York: Springer.
- Janßen, R., Tuerlinckx, F., Meulders, M., & De Boeck, P. (2000). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics*, 25, 285-306.
- Jonassen, D. H., Tessmer, M. & Hannum, W. H. (1999). *Task analysis methods for instructional design*. Mahwah, NJ: Erlbaum.
- Kaiser, G., Hino, K. & Knipping, C. (2005). Proposal for a framework to analyse mathematics education in Eastern and Western traditions – looking at England, France, Germany and Japan. In K. D. Graf & F. Leung (Eds.), *ICMI Comparative Study on Mathematics education in different cultural traditions: a comparative study of East Asia and the West*. Dordrecht, Kluwer Academics.
- Klauer, K. J. (1978). Perspektiven pädagogischer Diagnostik. In K. J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik* (S. 3-14). Düsseldorf: Schwann.
- Klauer, K. J. (1987). *Kriteriumsorientierte Tests*. Göttingen: Hogrefe.
- Klauer, K. J. (1994). Neuer Entwicklungen in der pädagogisch-psychologischen Diagnostik. *Diagnostica*, 40, 244-261.
- Klee, R., Sandmann, A. & Vogt, H. (Hrsg.) (2005). *Lehr- und Lernforschung in der Biologiedidaktik, Band 2*. Innsbruck: Studienverlag.
- Klieme, E. (2004). Assessment of cross-curricular problem-solving competencies. In J. H. Moskowitz & M. Stephens (Eds.), *Comparing learning outcomes. International assessments and education policy* (pp. 81-107). London: Routledge Falmer.
- Klieme, E. & Baumert, J. (2001). Identifying national cultures of mathematics education: Analysis of cognitive demands and differential item functioning in TIMSS. *European Journal of Psychology of Education*, 16, 383-400.
- Klieme, E., Funke, J., Leutner, D., Reimann, P., & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz? Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47, 179-200.
- Klieme, E., Leutner, D. & Wirth, J. (Hrsg.) (2005). *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 139-190). Opladen: Leske & Budrich.
- Kobayashi, M. (2002). Method effects on reading comprehension test performance: text organization and response format. *Language Testing* 19, 193-220.
- Köller, O. (2004). Schulische Leistungen am Ende der gymnasialen Oberstufe: Wichtige Ressourcen für den Übergang ins Studium und eine erfolgreiche Berufskarriere? *Zeitschrift für Erziehungswissenschaft*, 7, 3. Beiheft 2004, 185-199.
- Köller, O., Baumert, J. & Schnabel, K. U. (1999). Wege zur Hochschulreife: Offenheit des Systems und Sicherung vergleichbarer Standards. *Zeitschrift für Erziehungswissenschaft*, 2, 387-422.
- Köller, O., Watermann, R., Trautwein, U. & Lüdtke, O. (Hrsg.) (2004). *Wege zur Hochschulreife in Baden-Württemberg. TOSCA – Eine Untersuchung an allgemeinbildenden und beruflichen Gymnasien*. Opladen: Leske + Budrich.

- Kröner, S., Plass, J. L. & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence*, 33, 347-368.
- Kubinger, K. D. (1987). Adaptives Testen. *Tests und Trends*, 6, 103-127.
- Labudde, P., Herzog, W., Neuenschwander, M., Violi, E. & Gerber, C. (2000). Girls and physics: teaching and learning strategies tested by classroom interventions in grade 11. *International Journal of Science Education*, 22, 143-157.
- Leutner, D. (1992). Das Testlängendilemma in der lernprozessbegleitenden Wissensdiagnostik. *Zeitschrift für Pädagogische Psychologie*, 6, 233-238.
- Leutner, D. (2001a). Instruktionspsychologie. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 267-276). Weinheim: PVU.
- Leutner, D. (2001b). Pädagogisch-psychologische Diagnostik. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 521-530). Weinheim: PVU.
- Leutner, D. (2002). The fuzzy relationship of intelligence and problem solving in computer simulations. *Computers in Human Behavior*, 18, 685-697.
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147-175). Münster: Waxmann.
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2005). Die Problemlösekompetenz in den Ländern der Bundesrepublik Deutschland. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003: Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* (S. 125-146). Münster: Waxmann.
- Leutner, D. & Plass, J. L. (1998). Measuring learning styles with questionnaires versus direct observation of preferential choice behavior: Development of the Visualizer/Verbalizer Behavior Observation Scale (VV-BOS). *Computers in Human Behavior*, 14, 543-557.
- Leutner, D. & Wirth, J. (im Druck). What we have learned from PISA so far: A German educational psychology point of view. *Journal of Educational Policy*.
- van der Linden, W. (2005). A Comparison of Item-Selection Methods for Adaptive Tests with Content Constraints. *Journal of Educational Measurement*, 42, 283-302.
- van der Linden, W. & Hambleton, R. K. (1996). *Handbook of modern item-response theory*. Berlin: Springer.
- Lissmann, U. (2004). Beurteilung und Beurteilungsprobleme bei Portfolios. In R. S. Jäger (Hrsg.), *Von der Beobachtung zur Notengebung – ein Lehrbuch – Diagnostik und Benotung in der Aus-, Fort- und Weiterbildung* (S. 211-241). Landau: Verlag Empirische Pädagogik.
- Martin, R., Latour, T., Burton, R., Busana, G. & Vandenabeele, L. (2005). Covering different levels of evaluation needs by an internet-based computer-assisted testing framework for collaborative distributed test development and delivery. Full paper submitted for presentation at ED-MEDIA 2005 – World Conference on Educational Multimedia, Hypermedia & Telecommunications.
- McClelland, D. C. (1973). Testing for competence rather than for „intelligence“. *American Psychologist*, 28, 1-14.
- Möller, K. (2001). Wissenserwerb und Wissensqualität im naturwissenschaftsbezogenen Sachunterricht. In J. Kahlert & E. Inckemann (Hrsg.), *Wissen, Können und Verstehen – über die Herstellung ihrer Zusammenhänge im Sachunterricht* (S. 115-126). Bad Heilbrunn: Klinkhardt.
- Möller, J. & Schiefele, U. (2004). Motivationale Grundlagen der Lesekompetenz. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Entwicklung, Bedingungen und Förderung der Lesekompetenz: Vertiefende Analysen der PISA-2000-Daten* (S. 101-124). Wiesbaden: Verlag für Sozialwissenschaften.
- Molz, M., Eckhardt, A., Schnotz, W., Hochscheid-Mauel, D. & Niegemann, H. (2004). Deconstructing iInstructional design models: towards an integrative conceptual framework for in-

- structional design research. In H. Niegemann, R. Brünken & D. Leutner (Eds.), *Instructional design and multimedia learning*. Münster: Waxmann.
- Muthén, B. & Muthén, L. (2000). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24, 882-891.
- Neubauer, A. C., Grabner, R. H., Fink, A. & Neuper, C. (2005). Intelligence and neural efficiency: Further evidence of the influence of task content and sex on the brain-IQ relationship. *Cognitive Brain Research*, 25, 217-225.
- Neubauer, A. C., Grabner, R. H., Freudenthaler, H. H., Beckmann, J. F. & Guthke, J. (2004). Intelligence and individual differences in becoming neurally efficient. *Acta Psychologica*, 116, 55-74.
- Nold, G. (2003). DESI – a language assessment project in Germany and the pros and cons of large-scale testing. *Empirische Pädagogik* 17, 368-379.
- Nold, G. & Rossa, H. (im Druck a). Hörverstehen. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung in der DESI-Studie*. Weinheim: Beltz.
- Nold, G. & Rossa, H. (im Druck b). Leseverstehen. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung in der DESI-Studie*. Weinheim: Beltz.
- Nürk, H. C., Weger, U. & Willmes, K. (2005). Language effects in magnitude comparison: Small, but not irrelevant. *Brain and Language*, 92, 262-277.
- Oberauer, K., Schulze, R., Wilhelm, O. & Süß, H.-M. (2005). Working memory and intelligence – their correlation and their relation: A comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131, 61-65.
- Ordinate (2004). SET-10 Test Description & Validation Summary. Menlo Park, CA: Ordinate.
- Pellegrino, J. W., Chudowsky, N. & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: National Academic Press.
- Perels, F., Gurtler, T., Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction*, 15, 123-139.
- Plass, J. L., Chun, D., Mayer, R. E. & Leutner, D. (1998). Supporting visualizer and verbalizer learning preferences in a second language multimedia learning environment. *Journal of Educational Psychology*, 90, 25-36.
- Plichart, P., Jadoul, R., Vandenabeele, L. & Latour, T. (2004). TAO, a collaborative distributed computer-based assessment framework built on Semantic Web standards. Full paper presented at the International Conference on Advances in Intelligent Systems – Theory and Applications AISTA2004, in cooperation with IEEE Computer Society, 15-18 November 2004, Luxembourg.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U. (Hrsg.) (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J. & Schiefele, U. (Hrsg.) (2005). *PISA 2003. Der zweite Vergleich der Länder in Deutschland – was wissen und können Jugendliche?* Münster: Waxmann.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001). *Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse*. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 192-248). Opladen: Leske & Budrich.
- Reeff, J.-P. & Martin, R. (im Druck). Use of the internet for the assessment of students' achievement. In J. Hartig, E. Klieme & D. Leutner (Eds.): *Assessment of competencies in educational contexts: state of the art and future prospects*. Göttingen: Hogrefe & Huber.
- Reeff, J.-P., Zabala, A. & Klieme, E. (2004). ALL problem solving framework. In Statistics Canada (Ed.), *Adult Literacy and Life Skills Survey: frameworks, aspects of design, development and validation*. Ottawa: Statistics Canada.

- Reiss, K., Hellmich, F. & Thomas, J. (2002). Individuelle und schulische Bedingungsfaktoren für Argumentationen und Beweise im Mathematikunterricht. *Zeitschrift für Pädagogik*, 45. Beiheft, 51-64.
- Renkl, A. & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist*, 38, 15-22.
- Richter, T. & Christmann, U. (2002). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (S. 25-58). Weinheim: Juventa.
- Rost, J. (2003). Latent class analysis. In R. Fernandez-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (Vol. 1, p. 539-543). Thousand Oaks: Sage Publications.
- Rost, J., Carstensen, C. H., Bieber, G., Neubrand, M. & Prenzel, M. (2003). Naturwissenschaftliche Teilkompetenzen im Ländervergleich. In Deutsches PISA-Konsortium (Hrsg.): *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 109-129). Opladen: Leske und Budrich.
- Rost, J. & Carstensen, C. H. (2002). Multidimensional Rasch measurement via item component models and faceted designs. *Applied Psychological Measurement*, 26, 42-56.
- Rost, J. & von Davier, M. (1995). Mixture distribution Rasch models. In G. Fischer & I. Molenaar (eds.), *Rasch models: foundations, recent developments, and applications* (pp. 257-268). Berlin: Springer.
- Rost, J. & Langeheine, R. (Eds) (1997). *Applications of latent trait and latent class models in the social sciences*. New York, US: Waxmann Publishing Co.
- Rupp, A. (2005). Quantifying subpopulation differences for a lack of invariance using complex examinee profiles: an exploratory multigroup approach using functional data analysis. *Educational Research and Evaluation*, 11, 71-97.
- Rychen, D. S. & Salganik, L. H. (Eds.) (2001). *Defining and selecting key competencies*. Seattle: Hogrefe & Huber Publishers.
- Rychen, D. S. & Salganik, L. H. (Eds.) (2003). *Key competencies for a successful life and a well-functioning society*. Washington: Hogrefe & Huber Publishers.
- Schaffner, E., Schiefele, U., Drechsel, B. & Artelt, C. (2004). Lesekompetenz. In PISA-Konsortium Deutschland (Hrsg.). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 93-110). Münster: Waxmann.
- Schecker, H. (1999). TIMSS und die Folgen. In H. Günther-Arndt (Hrsg.), *Didaktische Forschung für die Schule* (S. 13-29). Universität Oldenburg: Zentrum für pädagogische Berufspraxis.
- Scheerens, J., Glas, C., Thomas, S. M. (2003). *Educational evaluation, assessment, and monitoring. A systematic approach*. Lisse: Swets and Zeitlinger.
- Schneider, W., Lockl, K., Fernandez, O. (2005). Interrelationships among theory of mind, executive control, language development, and working memory in young children: A longitudinal analysis. In W. Schneider, R. Schumann-Hengsteler, B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 259-284). Mahwah, NJ: Lawrence Erlbaum.
- Schnotz, W. & Dutke, S. (2004). Kognitionspsychologische Grundlagen der Lesekompetenz: Mehrebenenverarbeitung anhand multipler Informationsquellen. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 61-99). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schnotz, W., Vosniadou, S. & Carretero, M. (Eds.) (1999). *New perspectives on conceptual change*. Oxford: Elsevier.

- Schwippert, K., Bos, W. & Lankes, E.-M. (2004). Lesen Mädchen anders? Vertiefende Analysen zu Geschlechtsdifferenzen auf Basis der Internationalen Grundschul-Lese-Untersuchung IGLU. *Zeitschrift für Erziehungswissenschaft*, 7, 219-234.
- Simonton, K. (2003). Expertise, competence, and creative ability: The perplexing complexities. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 213-239). Cambridge: Cambridge University Press.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31, 15-21.
- Souvignier, E., Küppers, J. & Gold, A. (2003). Lesestrategien im Unterricht: Einführung eines Programms zur Förderung des Textverstehens in 5. Klassen. *Unterrichtswissenschaft*, 31, 166-183.
- Spiel, C. & Glück, J. (im Druck). A model based test of competence profile and competence level in deductive reasoning. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts: State of the art and future prospects*. Göttingen: Hogrefe.
- Spiel, C., Glück, J. & Gößler, H. (2001). Stability and change of unidimensionality: the sample case of deductive reasoning. *Journal of Adolescent Research*, 16, 150-168.
- Stanat, P., & Baumert, J. (Eds.) (2001). Large-scale assessment studies and their contributions to educational psychology (Special issue). *European Journal of Psychology of Education*, 16 (3).
- Stern, E. (in Druck). Neuroscience goes to school. *Science*.
- Sternberg, R. J. & Grigorenko, E. (Eds.) (2003). *The psychology of abilities, competencies, and expertise*. New York: Cambridge University Press.
- Steyer, R. (2005). Analyzing individual and average causal effects via structural equation models. *Methodology*, 1, 39-54.
- Sumfleth (2002). Empirical research in chemistry education in Germany – actual situation and perspectives in times of TIMSS and PISA. In B. Ralle & I. Eilks (Eds.), *Research in chemical education – What does this mean?* (pp. 57-68). Aachen: Shaker.
- Weinert, F. E. (2001a). Concept of competence: a conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Seattle: Hogrefe & Huber Publishers.
- Weinert, F. E. (Hrsg.). (2001b). *Leistungsmessung in Schulen*. Weinheim: Beltz.
- Weinert, F. E., Helmke, A. & Schrader, F.-W. (1992). Research on the model teacher and the teaching model: Theoretical contradiction or conglutination? In F. Oser, A. Dick & J. L. Patry (Hrsg.), *Effective and responsible teaching: The new synthesis* (S. 249-260). San Francisco: Jossey-Bass Publishers.
- Weinert, F. E & Schneider, W. (Eds.). (1995). *Memory performance and competencies: Issues in growth and development*. Hillsdale, NJ: Lawrence Erlbaum.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66, 297-333.
- Wild, K.-P. & Krapp, A. (2001). Pädagogisch-psychologische Diagnostik. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 515-563). Weinheim: Beltz PVU.
- Wilhelm, O., & Engle, R. (Eds.), (2005). *Understanding and measuring intelligence*. London: Sage.
- Willmes, K., Dambeck, N., Nuerk, H.-C. (2001). A fMRI study about notation- and task-specificity of functional activation for numbers. *Journal of the International Neuropsychological Society*, 7, 120-132.
- Wilson, M. (1989). Saltus: A psychometric model for discontinuity in cognitive development. *Psychological Bulletin*, 105, 276-289.
- Wilson, M. (2003). On choosing a model for measuring. *Methods of Psychological Research Online*, 8, 1-22.
- Wilson, M. (2005). *Constructing measures. An item response modelling approach*. Mahwah: Lawrence Erlbaum Associates.

- Wilson, M., de Boeck, P. & Carstensen, C. (im Druck). Explanatory item response models: A brief introduction. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts: state of the art and future prospects*. Göttingen: Hogrefe & Huber.
- Wirth, J. (2004). *Selbstregulation von Lernprozessen*. Münster: Waxmann.
- Wirth, J. & Klieme, E. (2003). Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy, & Practice*, 10, 329-345.
- Wottawa, H. (2001). Evaluation. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 152-158). Weinheim: Beltz PVU.

Anschrift der Autoren:

Prof. Dr. Eckhard Klieme, Deutsches Institut für Internationale Pädagogische Forschung,
Schloßstraße 29, 60486 Frankfurt a.M.

Prof. Dr. Detlev Leutner, Lehrstuhl für Lehr-Lernpsychologie, FB Bildungswissenschaften, Uni-
versität Duisburg-Essen, Campus Essen, 45117 Essen.